

ONZEminer

a digital tool for tracing the origins and
development of New Zealand English

Heidi Quinn

University of Canterbury

heidi.quinn@canterbury.ac.nz

Transformations'07, 27 August 2007

“Language is a central feature of human identity. When we hear someone speak, we immediately make guesses about gender, educational level, age, profession, and place of origin. Beyond this individual matter, a language is a powerful symbol of national and ethnic identity.”

Spolsky (1999: 181)

Present-Day New Zealand English (NZE) is characterised by a unique combination of phonological and grammatical features that make it easy for us to distinguish New Zealanders from speakers of other varieties.

cc-fyn94-12a, born 1972



I mean the guys are a lot better [...]

I think women are probably have to work at it a lot harder than y'know than the guys

cc-myn02-11a, born 1981



jumped out of the car and got the fire engine out and it turned out the fuel line fell off and put hot fuel on to the engine

The first generations of English-speaking New Zealanders sounded very different.

- lots of variation between speakers
- still strongly influenced by the varieties of English spoken by the early settlers, who mainly came from England, Scotland, and Ireland, often via Australia (cf. Gordon et al. 2004)

ia-98, born 1893, Otago, Irish parents

oh and when we were going to school
we used to go to the pictures [...]



I can remember the magic lantern shows
and we were very delighted about those

ia-70b, born 1896, Central Otago



and I don't know why I did it but I got sugar
and rubbed sugar into her hair to take the
stuff out of course . she was a terrible
mess

The ONZE Archives

- **Mobile Unit (mu)**

95 speakers born 1851-1919, recorded by the New Zealand Broadcasting service after WWII

- **Intermediate Archive (ia)**

about 90 speakers mostly born 1890-1935 and recorded by Rosemary Goodyear, Leslie Evans, and UC researchers in the early 1990s

- **Canterbury Corpus (cc)**

currently 407 speakers born 1926-1985, recorded by students at the University of Canterbury since 1994

ONZEminer

- a browser-based research tool developed by Jen Hay and Robert Fromont
- freely available to download from:
<http://www.ling.canterbury.ac.nz/jen/onzeminer>
- software currently integrated with ONZEminer
Transcriber (<http://trans.sourceforge.net/>)
Praat (<http://www.praat.org/>)
Stanford Parser
(<http://nlp.stanford.edu/downloads/lex-parser.shtml>)
CELEX (<http://www.ru.nl/celex/>)

ONZEminer

- ONZEminer is work in progress and its search and analysis facilities are constantly being refined by programmer Robert Fromont
- the program is already saving us hours of hand-searching and has allowed us to carry out linguistic comparisons on a much larger scale than was feasible before
- ONZEminer has also made it possible for undergraduate students to become involved in departmental research projects

projects using ONZEminer

- the ONZE project (cf. Gordon et al. 2007)
<http://www.ling.canterbury.ac.nz/onze/index.html>
- MAONZE
<http://www.ece.auckland.ac.nz/~cwat057/MAONZE/MAONZE.html>
- a project on Maori English by Margaret Maclagan and Jeanette King
- Katie Drager's PhD research on language and social identity (cf. Drager 2006)
- Alex D'Arcy's project on regional variation in Canterbury

projects using ONZEminer

- collaborative work on *give* with Stanford University (cf. Hay & Bresnan 2006; Bresnan & Hay in press)
- joint projects with Newcastle University (UK) on phrase-final /t/ (cf. Docherty et al. 2006) and on quotatives (current work by Alex D'Arcy and Isa Buchstaller)
- interactive corpus of German dialects (currently being created by Heidi Quinn)

http://localhost:8080/miner/

Google

Go

Bookmarks

5 blocked

Check

AutoLink

AutoFill

Send to

Settings

del.icio.us

TAG

onze miner

Home

Page

Tools

ne - filter - upload - speakers - transcripts - utilities - transcript families - corpora - transcript types - field names - class options - layers - celex layers - activity

onze miner

- install required components for sound to work
- filter speakers
- credits

Version 20061204.1602

Copyright © 2004-2006 **ONZE Project** University of Canterbury, NZ

This is free software distributed under the terms of the **GNU General Public License**

http://localhost:8080/miner/speakers.jsp

Google

Go

Bookmarks

5 blocked

Check

AutoLink

AutoFill

Send to

Settings

Transcript - Speaker List

Home

Page

Tools

ne - filter - upload - speakers - transcripts - utilities - transcript families - corpora - transcript types - field names - class options - layers - celex layers - activity - ?

speakers

	name	transcripts	corpus	gender	yob	class	region	origin mother	origin father
<input type="checkbox"/>	Family:	<input type="text"/>	CC	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
					to				
<input checked="" type="checkbox"/>	fyn94-12a	6	CC	F	1972	N			
<input type="checkbox"/>	Interviewer	46	ME	null	null	null	null	null	null
<input checked="" type="checkbox"/>	MYN02-11a	5	CC	M	1981	N	null	null	null
<input type="checkbox"/>	not transcribed	2	CC	null	null	null	null	null	null
<input type="button" value="search"/> <input type="button" value="layered search"/>									

Version 20061204.1602

Copyright © 2004-2006 ONZE Project University of Canterbury, NZ

This is free software distributed under the terms of the GNU General Public License

search

name	gender	class	year of birth	
fyn94-12a	F	N	1972	[notes]
MYN02-11a	M	N	1981	[notes]

Text/Regular Expression:

☒ Only search transcripts for which these are the main speakers.


☐ Export results to Excel (CSV) [options]



[transcript types]


search

results


Found 8 results

fyn94-12a-08.trs - **fyn94-12a:**

1. ☐   don't think oh it didn't rain I'm not sure why they didn't actually .

fyn94-12a-09.trs - **fyn94-12a:**

2. no results

fyn94-12a-10.trs - **fyn94-12a:**

3. ☐   you think like that that's why I li

search

name	gender	class	year of birth	
fyn94-12a	F	N	1972	[notes]
MYN02-11a	M	N	1981	[notes]

Text/Regular Expression:

- ☒ Only search transcripts for which these are the main speakers.
- ☒ Export results to Excel (CSV) [options]

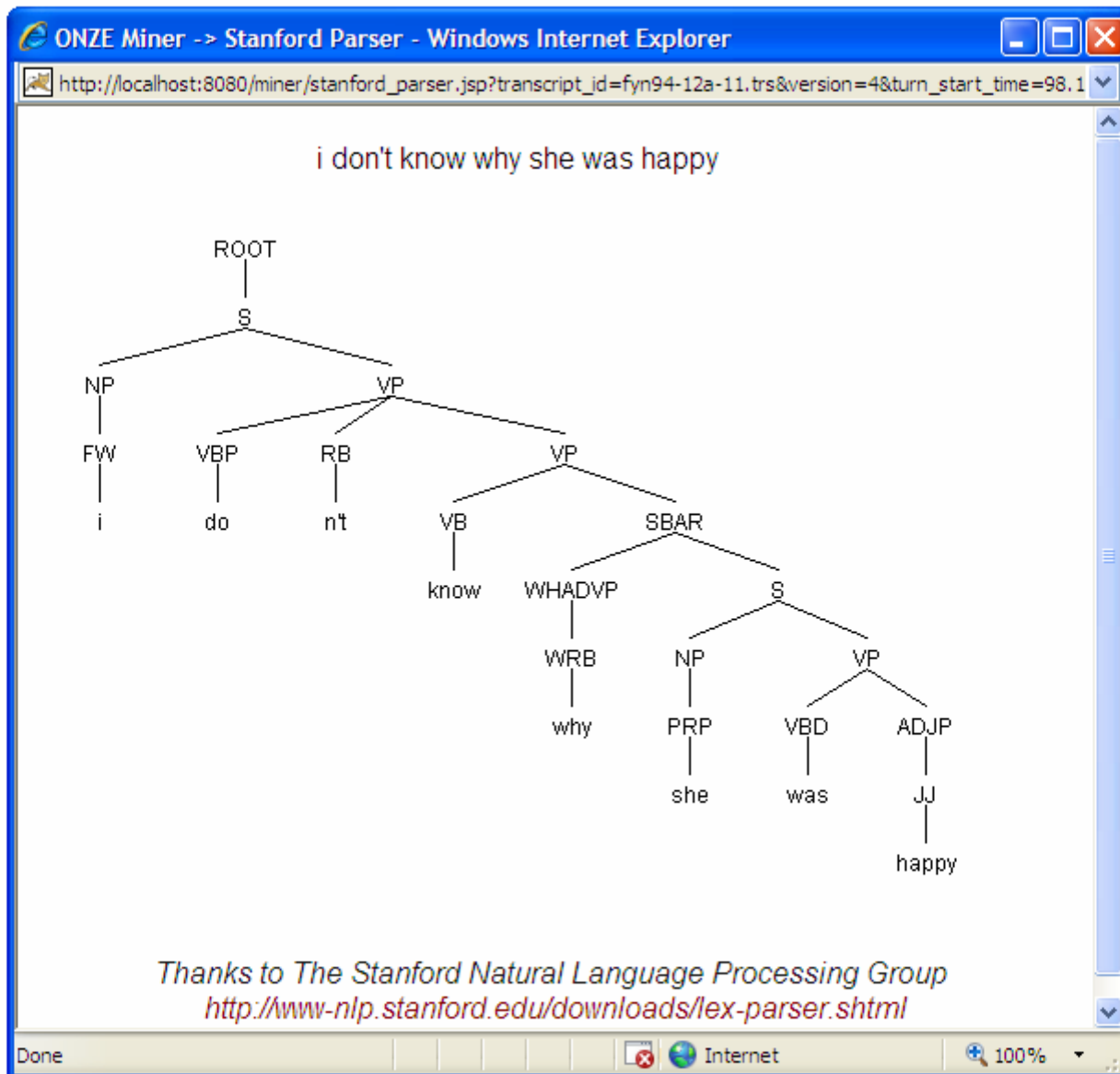
[transcript types]

ft Excel - results_why[1].csv

Edit View Insert Format Tools Data Window Help

Type a question for help

	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Number	Transcript	Speaker	Gender	YearOfBirth	class	Sync	URL	Text						
1	fyn94-12a	fyn94-12a	F	1972	N	271.889	http://local	I don't think oh it didn't rain I'm not sure why they didn't actually .						
2	fyn94-12a	fyn94-12a	F	1972	N	285.174	http://local	you think like that that's why I li						
3	fyn94-12a	fyn94-12a	F	1972	N	119.929	http://local	I don't know why . she was happy						
4	fyn94-12a	fyn94-12a	F	1972	N	262.94	http://local	I mean I had a awful birth that's probably why I think that						
5	fyn94-12a	fyn94-12a	F	1972	N	4.254	http://local	but it's why take that risk though it's you and a baby .						
6	myn02-11a	MYN02-11	M	1981	N	36.896	http://local	quite understand why it stalled						
7	myn02-11a	MYN02-11	M	1981	N	175.09	http://local	yeah . why ?						
8	myn02-11a	MYN02-11	M	1981	N	176.28	http://local	why would you want a Holden ?						



it-to NLP hoping that it would be a girl

it-to NLP cos I was thinking oh it could be a girl .

filter - upload - speakers - transcripts - utilities - transcript families - corpora - transcript types - field names - class options - layers - c
 layers - activity - ?

layered search

speakers

name	gender	class	year of birth	
fyn94-12a	F	N	1972	[notes]
MYN02-11a	M	N	1981	[notes]

layers

- ☐ ab morphology
☐ ab syntax
☒ ø₃ phonology
☐ orthography
☒ transcript

search across words.

search regular expressions

phonology « followed by
 followed by **transcript**

☒ Only search transcripts for which these are the main speakers.

☐ Only show results from the first transcripts.

[transcript types]

Tramyn02-11a-WL-08.trs

next>>

am Files\Apache Software Foundation\Tomcat 5.5\webapps\miner\files\CC\myn02-11a-WL\wav\myn02-11a-WL-08.tr
Message sent to Praat

speakers

layers

id	name	check	type	dialect	accent	scope
spk1	myn02-11a	no		native		local

☐ ab morphology

☐ ab syntax

☒ ø phonology

☒ Use IPA

☐ Use SIL IPA Font

☐ Use ASCII representation

☒ transcript

display layers

transcript

myn02-11a

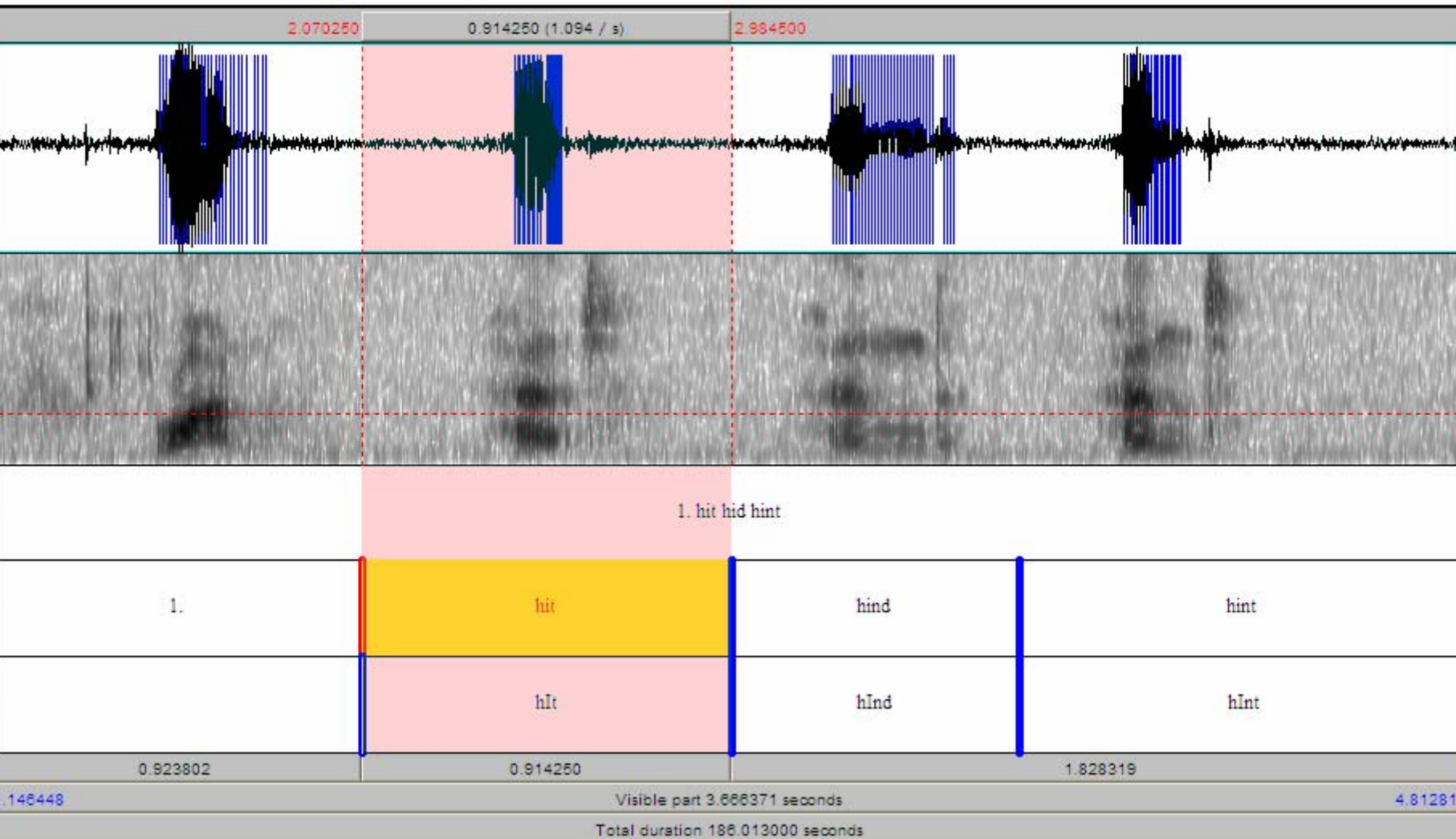


Wordlist



hit hid hint

1. hit hid hint



changes involving consonants

- loss of rhoticity and rise of linking & intrusive /r/ (cf. Hay & Sudbury 2005)

ia-98

cc-myn02-11a

magic lantern

turned

fire engine



- loss of /hw/ (cf. Schreier et al. 2003)

why

ia-70b



cc-fyn94-12a



changes involving vowels

- closing diphthongs
(cf. Maclagan et al. 1999; Hay et al. in press)

now ia-98  cc-fyn94-12a 

- the NEAR/SQUARE merger (cf. Maclagan & Gordon 1996; Gordon & Maclagan 2001, Langstrof 2004)

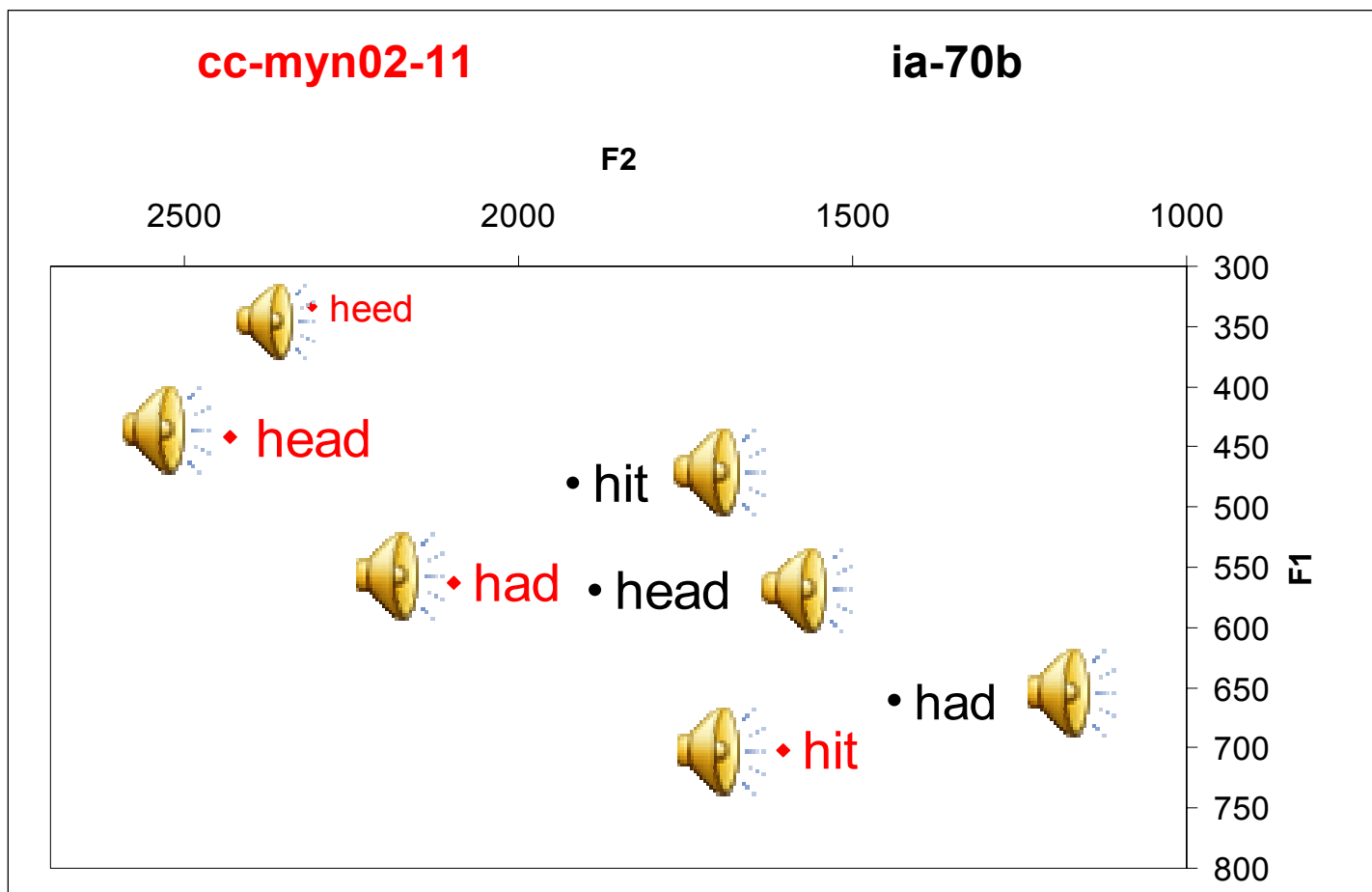
here hair ia-98  cc-fyn94-12a 

changes involving vowels

- raising of TRAP and DRESS (Maclagan et al. 1999; Gordon et al. 2004; Maclagan & Hay 2007)
- centralisation of KIT (Gordon et al. 2004; Langstrof 2006)
- diphthongisation of FLEECE (Maclagan & Hay 2007)

The changes to TRAP, DRESS and KIT are illustrated in the vowel plot on the next slide.

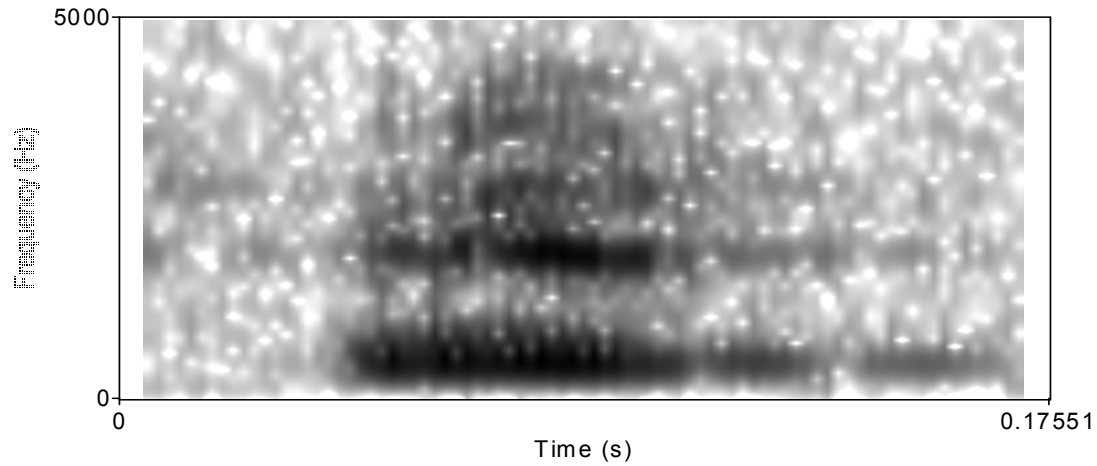
Changes in NZE front vowels



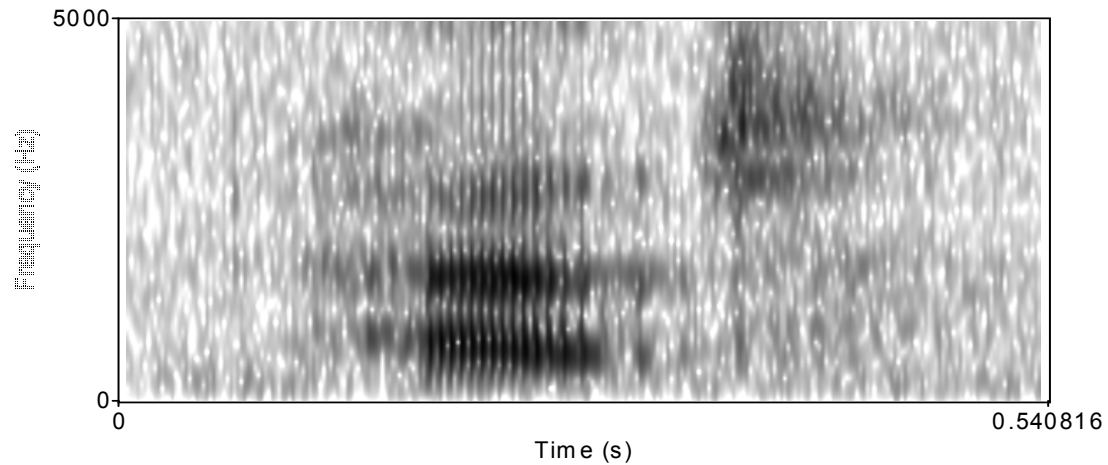
note: The examples for the younger speaker come from a wordlist, whereas the tokens for the older speaker were produced in connected speech. This exaggerates the differences between their vowels.

Spectrograms for 'hit'

ia-70b 'hit'



cc-myn02-11a 'hit'



grammatical changes

- subject-verb agreement with *be* (cf. Hay & Schreier 2004)

the volunteers was getting a little hard pressed (mu87a, male, born 1888)

there was just heaps of shops 'n you know both sides of the road (cc-fyn94-12a)

grammatical changes

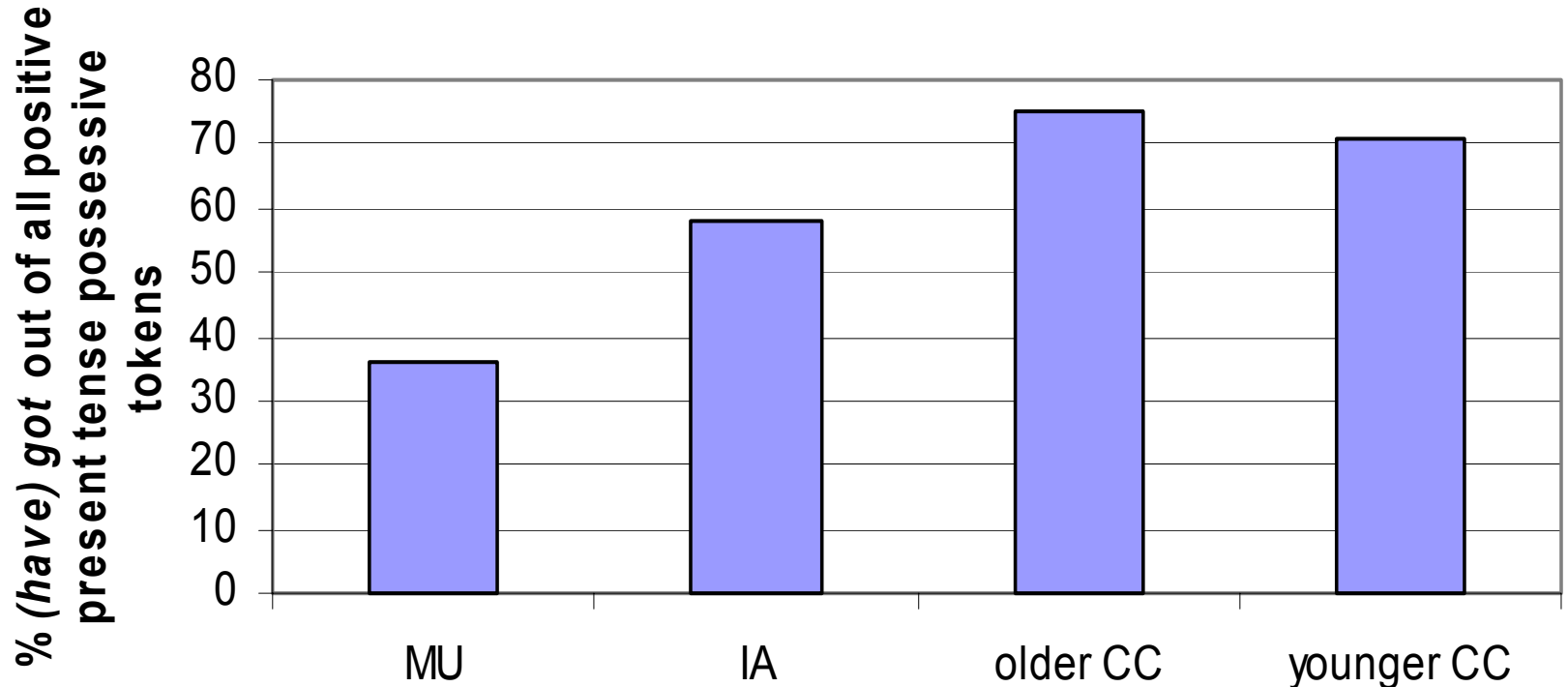
- the use of (HAVE) *got* in present tense possessives (cf. Quinn 2004 & to appear)

now I **have** two young children (ia-70b)

I'**ve got** a soft spot for them (cc-myn02-11a)

he'**s got** the size for hockey (cc-fyn94-12a)

(HAVE) *got* by corpus



older CC = aged 45-60 when interviewed
younger CC = aged 20-30 when interviewed

the future of NZE

- in the 2006 census, 14.6% of New Zealanders identified themselves as Maori, and 7% as Pacific Islanders
- more than a third of the Maori and Pacific Island population were under the age of 15

(Statistics New Zealand 2007)

the future of NZE

- young New Zealanders with Maori and Pacific Island backgrounds are increasingly using English to express their ethnic identity (cf. Taumoefolau et al. 2002; Starks & Reffell 2005 & 2006; Kennedy 2006; Benton 2007; Szakay 2006 & 2007; Hay et al. in press; Bell & Gibson forthcoming)
- you don't have to be of Maori or Pacific Island descent to speak one of these new varieties of English

- **speaker 1**



cos I mean in recent years yous have had a pretty s pretty solid score specially in prems like yous have done pretty good

- **speaker 2**



eye for an eye tooth for a tooth [...]
there's a natural sort of connection with the church family

- **speaker 3**



it sucks getting the sports results before you even see the game [...] always you can't miss it when it's on the front page

note: all three speakers come from Margaret MacLagan and Jeanette King's Maori English corpus

Acknowledgements

Special thanks to:

- Robert Fromont, Jen Hay, and Katie Drager for help with ONZEminer, Praat, and vowel analysis
- Margaret Maclagan and Jeanette King for giving me access to their Maori English corpus
- Liz Gordon for drawing my attention to recent demographic data and for help with selecting speakers to demonstrate ONZEminer on
- Alex D'Arcy, Kate Kearns, and Andrew Carstairs-McCarthy for valuable comments on an earlier version of this talk

References

- Bell, Allan and Andy Gibson. forthcoming. Pasifika English in New Zealand. Paper to be presented at NWAV 36, Philadelphia, 11-14 October 2007.
- Benton, Richard A. 2007. *Mauri* or mirage? The status of the Maori language in Aotearoa New Zealand in the third millennium. *Language policy, culture, and identity in Asian contexts*, ed. by Amy B.M. Tsui and James W. Tollefson, 163-181. Mahwah, NJ: Lawrence Erlbaum Associates.
- Bresnan, Joan and Jennifer Hay. in press. Gradient grammar: an effect of animacy on the syntax of 'give' in varieties of English. *Lingua*.
- Docherty, Gerry, Jennifer Hay, and Abby Walker. 2006. Sociophonetic patterning of phrase-final /t/ in New Zealand English. *Proceedings of the 11th Australian International Conference on Speech Science & Technology*, ed. by Paul Warren and Catherine I. Watson, 378-383. Australian Speech Science & Technology Association.
- Drager, Katie. 2006. Social categories, grammatical categories, and the likelihood of “like” monophthongisation. *Proceedings of the 11th Australasian International Conference on Speech Science & Technology*, ed. by Paul Warren and Catherine I. Watson, 384-387. Australian Speech Science & Technology Association.

- Gordon, Elizabeth, Lyle Campbell, Jennifer Hay, Margaret MacLagan, Andrea Sudbury, and Peter Trudgill. 2004. *New Zealand English: its origins and evolution*. Cambridge: Cambridge University Press.
- Gordon, Elizabeth and Margaret MacLagan. 2001. 'Capturing a sound change': a real time study over 15 years of the NEAR/SQUARE diphthong merger in New Zealand English. *Australian Journal of Linguistics* 21. 215-238.
- Gordon, Elizabeth, Margaret MacLagan, and Jennifer Hay. 2007. The ONZE corpus. *Models and methods in the handling of unconventional digital corpora, volume 2: diachronic corpora*, ed. by Joan C. Beal, Karen P. Corrigan, and Hermann L. Moisl. Palgrave.
- Hay, Jennifer and Joan Bresnan. 2006. Spoken syntax: the phonetics of giving a hand in New Zealand English. *The Linguistic Review* 23. 321-349.
- Hay, Jennifer, Margaret MacLagan, and Elizabeth Gordon. in press. *New Zealand English*. Edinburgh: Edinburgh University Press.
- Hay, Jennifer and Daniel Schreier. 2004. Reversing the trajectory of language change: subject verb agreement with BE in New Zealand English. *Language Variation and Change* 16. 209-235.
- Hay, Jennifer and Andrea Sudbury. 2005. How rhoticity became /r/-sandhi. *Language* 81. 799-823.
- Kennedy, Marianna. 2006. Variation in the pronunciation of English by New Zealand school children. PhD thesis, Victoria University of Wellington.

- Langstrof, Christian. 2004. The centring diphthongs of New Zealand English in the Intermediate Period: an acoustic analysis. *Proceedings of the 10th Australian International Conference on Speech Science & Technology*, 207-212. Australian Speech Science & Technology Association.
- Langstrof, Christian. 2006. Developments in New Zealand English in the Intermediate Period. PhD thesis, University of Canterbury, NZ.
- Maclagan, Margaret A. and Elizabeth Gordon. 1996. “Out of the AIR and into the EAR”: another view of the New Zealand diphthong merger. *Language Variation and Change* 8: 125-147.
- Maclagan, Margaret A., Elizabeth Gordon, and Gillian Lewis. 1999. Women and sound change: conservative and innovative behaviour by the same speakers. *Language Variation and Change* 11. 19-41.
- Maclagan, Margaret and Jennifer Hay. 2007. Getting fed up with our feet: contrast maintenance and the New Zealand English “short” front vowel shift. *Language Variation and Change* 19. 1-25.
- Quinn, Heidi. 2004. Possessive *have* and (*have*) *got* in New Zealand English. Paper presented at NWAV 33 in Ann Arbor, 1 October 2004.
- Quinn, Heidi. to appear. Alienability and stative HAVE in New Zealand English. *Proceedings of the CRLC Workshop on Grammatical Change, Brisbane 2006*, ed. by Rachel Hendery and Jennifer Hendriks.

- Schreier, Daniel, Elizabeth Gordon, Jennifer Hay, and Margaret Maclagan. 2004. The regional and linguistic dimension of /hw/ maintenance and loss in early 20th century New Zealand English. *English World-Wide* 24. 245-270.
- Spolsky, Bernard. 1999. Second-language learning. *Handbook of language and ethnic identity*, ed. by Joshua A. Fishman, 181-192. Oxford: Oxford University Press.
- Starks, Donna and Hayley Reffell. 2005. Pronouncing your Rs in New Zealand English? A study of Pasifika and Maori students. *New Zealand English Journal* 19. 36-48.
- Starks, Donna and Hayley Reffell. 2006. Reading 'TH': vernacular variants in Pasifika Englishes in South Auckland. *Journal of Sociolinguistics* 10. 382-392.
- Statistics New Zealand. 2007. QuickStats about Maori. Revised 27 March 2007.
- Szakay, Anita. 2006. Rhythm and pitch as markers of ethnicity in New Zealand English. *Proceedings of the 11th Australasian International Conference on Speech Science & Technology*, ed. by Paul Warren and Catherine I. Watson, 421-426. Australian Speech Science & Technology Association.
- Szakay, Anita. 2007. Identifying Maori English and Pakeha English from suprasegmental cues: a study based on speech resynthesis. MA thesis, University of Canterbury, NZ.
- Taumoe folau, Melenaitē, Donna Starks, Karen Davis, and Allan Bell. 2002. Linguists and language maintenance: Pasifika languages in Manukau, New Zealand. *Oceanic Linguistics* 41. 15-27.